DOCUMENT RESUME

ED 435 386 IR 057 536

AUTHOR Tsai, Shu-En
TITLE Comparison of the Effect of Using Title Keyword Searching and Subject Headings among the 10 Divisions of Dewey Decimal Classification.
PUB DATE 1998-12-00
NOTE 44p.; Master's Research Paper, Kent State University.
PUB TYPE Dissertations/Theses (040)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Comparative Analysis; *Dewey Decimal Classification; *Information Retrieval; *Keywords; *Search Strategies; *Subject Index Terms; Tables (Data)
IDENTIFIERS *Library of Congress Subject Headings; *Titles

ABSTRACT

Bibliographic records taken from books listed in "OCLC Selected Titles for Research and University Libraries" are used to determine whether the use of terms in the title for subject searching is an effective alternative to the use of Library of Congress subject headings among the 10 divisions of Dewey Decimal Classification. Terms in each title are tested with term(s) in the first element of every Library of Congress subject heading. Three hypotheses are tested: (1) sciences and technology subject areas have the highest match rate; (2) match rate in the social sciences is much lower than that of sciences and technology subject areas; and (3) title keyword is an effective alternative to subject headings in sciences and technology subject areas. Among the 10 Dewey divisions, the 500 division, natural sciences and mathematics, has the highest subject heading exact match, a rate of 56.2% in this study. The 800 division, disciplines in literature and rhetoric, accounts for the lowest percentage of subject exact match, 19.04%. (Author/MES)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 435 386

COMPARISON OF THE EFFECT OF USING TITLE KEYWORD SEARCHING AND SUBJECT HEADINGS AMONG THE 10 DIVISIONS OF DEWEY DECIMAL CLASSIFICATION

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
HAS BEEN GRANTED BY

D.P. Wallace

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

A Master's Research Paper submitted to the
Kent State University School of Library Science
in partial fulfillment of the requirements
for the degree Master of Library Science

by

Shu-En Tsai

December, 1998

IR057536

ABSTRACT

Bibliographic records taken from books listed in OCLC Selected Titles for Research and University Libraries are used to determine whether the use of terms in title as subject searching is an effective alternative to the use of Library of Congress subject headings among the 10 divisions of Dewey Decimal Classification. Terms in each title are tested with term(s) in the first element of every Library of Congress subject heading. Three hypotheses are tested: 1) sciences and technology subject areas have the highest match rate; 2) match rate in social sciences is much lower than that of sciences and technology subject areas; 3) title keyword is an effective alternative to subject heading in sciences and technology subject areas. Among the 10 Dewey divisions, the 500 natural sciences and mathematics division has the highest subject heading exact match, a rate of 56.20% in the study. In the 800 division, disciplines in literature and rhetoric, accounts for the lowest percentage of subject exact match, 19.04%.

Master's Research Paper by

Shu-En Tsai

M.A., Ohio State University, 1990

M.L.S., Kent State University, 1998

Approved by

Adviser Tschera H. Connell Date Dec. 8, 1998

v

TABLE OF CONTENTS

ACKNOWLEDGEMENTS viii

INTRODUCTION 1

LITERATURE REVIEW 2

Research Concluding that Title Keywords and Subject Headings Complement Each Other in Searching

Research Concluding that Title Keywords Outperform Subject Headings in the Volume of Output

Research Concluding that LCSH is Indispensable

Research Pertinent to System Design as an Approach to Improve Subject Access

Research Looking at Subject Access Analysis in Different Disciplines

PROBLEM STATEMENT AND OBJECTIVES OF THE RESEARCH 16

METHODOLOGY. 19

RESULTS. 25

DISCUSSION. 30

CONCLUSION. 34

REFERENCE LIST. 37

ACKNOWLEDGEMENTS

I am indebted to my advisor Dr. Tschera Connell who has graciously and tirelessly read through this paper several times, given me invaluable suggestions to improve the quality of this paper, and even taken time to correct my grammatical errors. Without her advice and assistance, I could not have presented this paper as it is.

I. INTRODUCTION

This study considers whether title keyword searching is an alternative to controlled vocabulary searching using subject headings.

When cataloging library materials, catalogers assign one or more subject headings to each catalog record to enhance the accessibility of the record. With the assistance of high technology, information users are now able to retrieve records by keywords. In light of this, one may argue that the term used in subject headings usually appears in the title field, and therefore, subject headings are redundant and thus not needed. Besides, it is costly and time-consuming to assign subject headings.

However, on the other hand, some researchers believe that subject headings continue to provide invaluable access points for retrieval of catalog records (e.g. Gerhan 1989; Carlyle 1989; Marner 1993). Title keyword searching is no substitute for retrieval by subject headings as the term of subject headings is not usually in the title field. Besides, subject headings lead users to other related terms, and group together the materials that are on the same subject but have different titles. Title keyword searching does not have these features.

Keyword searching in both title and subject fields is a standard feature of online catalogs today. With the advancement of technology, the user interface has improved a great deal. Online systems have enhanced the number of access points to increase recall. While many articles in library literature suggest a decline in subject searching and an increase in title keyword searching, many researches also show increasing concerns on the relevancy of subject access and information overflow. The issues of subject access today

are still the same as they were decades ago. Users, using title keyword searching strategy, have complained about retrieving too large result sets and yet still can not find enough information on the topic they look for. Title keyword searching does help improve recall but usually ends up with poor precision. As a result, library professionals are still debating the advantages and disadvantages of keyword free text searching versus controlled vocabulary searching.

II. LITERATURE REVIEW

Thomas A. Peters and Martin Kurth (1991) analyzed transaction logs of dial access search sessions from the online catalog of the University of Missouri. The objective of the study is to determine situational characteristics, and to examine how patrons use the combination of controlled and uncontrolled vocabulary in subject searching in an academic library online catalog. They selected the transactions in which both controlled (subject term(s)) and uncontrolled (title keyword) vocabulary keyword searches were conducted in the same search session. They found that nearly 59% of the mixed access subject search sessions started out as an uncontrolled vocabulary search. The study further indicated that users tended to stay in a session longer (user persistence) during controlled vocabulary leg (average 2.74 searches) when compared with uncontrolled vocabulary leg (average 1.93 searches). In terms of search output, each controlled vocabulary leg retrieved 19.23 bibliographic records on the average, while uncontrolled vocabulary leg retrieved 105.25 bibliographic records.

Peters and Kurth made the following four suggestions for the use of title keywords:

1. Users come to the search session with at least one known item. The user uses a title keyword search to retrieve the bibliographic record for the known item, note the assigned subject headings and execute a controlled vocabulary subject search on the most promising assigned heading to identify and locate other items that have been assigned the same heading; (Peters and Kurth, 210)

2. Users also use a known item and the title keyword search to achieve a simultaneous synthesis between known item and subject searching by carefully choosing the keywords from the title of the known item that are likely to retrieve other items of interest; (Peters and Kurth, 210)

3. Users try the title keyword search as uncontrolled vocabulary subject search simply to retrieve at least one potentially pertinent item. If such an item is retrieved, the assigned subject headings can be examined. Controlled vocabulary subject searches can then be used to increase the recall of pertinent items; (Peters and Kurth, 210)

4. Title keyword searches can be used for subject access in lieu of the controlled vocabulary subject searching. (Peters and Kurth, 210)

Peters and Kurth's findings show that users use both uncontrolled vocabulary subject searching and controlled vocabulary subject searching in a search session, which indicates that both subject searching methods complement each other.

Research Concluding that Title Keywords and Subject Headings Complement Each Other in Searching

Joy Tillotson (1995) examined three aspects of keyword searching to determine if keyword searches could be considered as a solution to problems of subject searching that patrons experience in online catalogs. The study also looked into the usefulness of the search result sets when patrons conducted keyword searching. Instead of analyzing transaction logs, Joy Tillotson conducted a keyword search study by carrying out 400 subject searches in two online catalogs of different sizes: one of about 700,000 records and the other of about 7 million records. The study found that keyword searching retrieved more items (with useful citations) than did controlled vocabulary searching. She matched keyword searches with a set of relevant materials, which was created by performing subject searches using terms that matched or closest matched LCSH. The average recall of relevant

materials using keyword searching was 68% in the larger database, and 73% in the smaller database.

Tillotson further conducted a survey via interviews to determine the level of user satisfaction in using title keyword versus controlled vocabulary in searching. The findings of this survey, however, revealed that "some keyword searches provided citations that appeared to be about the topic but were still declared unsuccessful by the searcher" (Tillotson, 203). Tillotson thus concluded that both keyword searching and controlled vocabulary searching should co-exist in an online catalog.

Hong Xu and Lancaster (1998) conducted research to investigate to what extent subject access points, which are available in titles and classification numbers, are not already being provided in subject headings in a common cataloging record. Hong Xu and Lancaster analyzed 205 items randomly selected from WorldCat (the Online Computer Library Center Online Union Catalog). These items were selected from the materials classified in Dewey Classification classes 300, 500, 600, and 700. Xu then assigned 844 unique subject access points (SAPs) to these 205 items, resulting in 4.11 SAPs per item. The term "subject access point" was defined as "any element in a bibliographic record that is indicative of the subject of the item represented, such as subject headings, a classification number, or words in titles, or subject headings." (Xu and Lancaster, 61)

Among the 205 items analyzed, a total of 634 SAPs were found in Subject Headings, averaging 3.09 SAPs in Subject Headings per item. A total of 458 SAPs were found in titles that averaged 2.23 SAPs in titles per item. A total of 406 SAPs were found in classification numbers, averaging 1.98 SAPs per item. Of the SAPs analyzed 328 SAPs were found to be duplicated in subject headings and titles; 222 SAPs were duplicated in titles and classification numbers; and 210 SAPs overlapped in all three categories. In other words, there was a 30.03% overlap among the SAPs in both titles and subject headings. About 32.96% of the SAPs in subject headings were not available in the other two

categories, and about 25.76% of SAPs in titles were not available in the other two categories. The findings of this research show that both subject headings and title keywords complement each other in improving subject access.

Research Concluding that Title Keywords Outperform Subject Headings in the Volume of Output

Some studies demonstrated that the number of results from title keyword searching in online catalogs is higher than that of subject heading searching. John Akeroyd (1990) conducted research attempting to test effectiveness in information seeking, and to infer a body of evidence on the ways three online catalogs were being used. He used transaction logs to evaluate information retrieval of three different interfaces of three online catalogs. The three systems were GEAC, Dynix, and LIBERTAS systems.

Geac, at the time, offered an unusual feature in subject searching. When patrons entered a subject search query, the system displayed a list of subject headings which closely matched the search query. When patrons made a selection from the list, the system linked the patrons to a classified sequence which enables patrons to browse backwards or forwards. In other words, the system design did not allow patrons to go from a subject search query directly to a bibliographic record, unless it was an exact match between the search query and the unique classification number. When testing subject searching in the GEAC system, Akeroyd verified that "searching for subject within title search was common practice" (Akeroyd, 38). GEAC's transaction log revealed 55% of the searches were title searches and 12% were subject searches. Comparing Dynix with GEAC, he commented "the intelligent application of title keyword searching was able to retrieve a corpus of relevant documents to most subject queries" (Akeroyd, 40). However, Akeroyd did not compare the relevancy of the output from either title keyword searching or subject heading searching.

Another similar report is from Ray R. Larson (1991). Larson analyzed data collected via transactions logs over a six-year period to determine the long-term trends, patterns of subject searching, and the changes in index usage in an online catalog. He gathered data by using search commands and analyzing the results for subject search frequency and title keyword search frequency. The definition of a "subject search" refers to a command mode search using either the SU (subject keyword, 600 field) or the XS (exact subject, 600 field) indexes as the index. A "title keyword search" refers to a command mode search using the TW (title words) index or the TI (exact title) index.

The results of his research showed a graduate but constant decline in subject searching – 0.0059% per day, and a slow increase in the use of title keyword searching – an average of 0.0077% per day. Larson's further analysis suggests that it was due to users' frustration in subject searching, especially search failure using Library of Congress subject heading. The switch to title keyword searching, however, placed great burden on users in finding synonyms to the search terms. On the other hand, according to Larson's analysis on the mean number of items retrieved using keyword indexes, title keyword searches usually retrieved a much more manageable size of result set than did subject searches.

Pat Ensor (1992) conducted a survey attempting to determine which patrons use keyword searching, and how keyword searching was being used. Ensor conducted the survey to gather information on patrons' use of keyword searching feature available on the NOTIS online catalog. The system supports Boolean operators. Ensor's findings showed that keyword searchers did more (42.6%) on "topic words" searching (searching words that were not necessarily Library of Congress subject headings) than they did Library of Congress subject heading searching (15.7%). Ensor did not clearly define which index fields are included in the "topic word" searching.

Joan M. Cherry (1989) took a similar approach. She collected data by observing and recording searchers' search sessions, and by asking the searchers to complete a questionnaire. Her study was designed to determine solutions to searches with zero-hits. 19 types of searches were performed online on 42 zero-hit subject searches in an attempt to prove that converting these search queries into other search forms, such as keyword title search, truncated original query, and word pairs from the original query, would improve recall.

Cherry reports that 62% of the hits resulting from keyword title, 43% of the hits resulting from title searches, and 33% of the hits resulting from keyword subject are useful. However, only 33% of the hits resulting from subject searches on LCSH on CD-ROM are useful. Cherry's findings indicate that, for a large zero-hit set, the better choice to improve recall is to convert the original queries into a keyword subject, keyword title, or title search. Her findings also led her to conclude that "educating users in the use of LCSH or cross-references will not solve the problems with the majority of zero-hit subject searches" (Cherry, 99).

Research Concluding that LCSH is Indispensable

Some studies, on the other hand, report that Library of Congress subject heading searching still plays an important role in subject access. An example is a study conducted by David R. Gerhan (1989). He studied the terms used in both title and subject heading fields to compare the effectiveness of title field keyword and subject heading field subject access in online catalogs. Gerhan randomly drew 391 sample bibliographic records from card catalog records of the Union College Library. Each card contained title and all assigned Library of Congress subject headings. Gerhan himself examined each card and made a judgement about the "usefulness" of terms in title and subject heading fields, with the

assumption that users would consult with a reference librarian, and turn to LCSH for "see" and "see also" for references.

His findings demonstrate that 76% of the sample would offer some degree of subject searching through title field in an online cataloging environment (provided that component of title words can be searched). Gerhan commented that "observable in this cohort of record is a sizable number that would offer only weak access" (Gerhan, 85). He identified 175 records out of the 391 sample records that contained title words that are "only slightly descriptive because of obscure, ambiguous, or obsolete wording"(page 85). In other words, the title words in the 175 records would have to be enhanced to achieve successful subject retrieval.

Gerhan also used his professional judgement to compare the performance of Library of Congress subject headings in enhancing subject searching. His findings showed that subject enhancement of the assigned Library of Congress subject headings was beneficial to 43% of the sample items. 24% of the sample indicated that both Library of Congress subject headings and titles combinations were needed in subject access. 5% of the items showed that terms from Library of Congress subject headings were indispensable.

As a result, Gerhan concluded that "Library of Congress subject headings and title field subject retrieval in an online setting, may be complementary, enhancing each other by providing routes around each other's weakness" (Gerhan, 87). His findings suggest that "Library of Congress subject heading is likely to provide the more effective subject access four times as often as will title keywords" (Gerhan, 87).

Another similar report is from Allyson Carlyle (1989). Carlyle used a list of matching categories to measure to what extent subject searching language of users would match LCSH. She used the transaction logs from ORION, the UCLA Library's Online Information System to study the matching of user expressions with LCSH. She defined the following three matching categories -- 1. Single heading match including both the exact match and

partial match; 2. Multiple headings match also including both the exact match and partial match; and 3. No match.

Taking every tenth subject search statement, Carlyle collected 171 user expressions. She then searched each expression to see if there was any single heading match against the subject fields of ORION. When no match occurred, she searched it against hard copy LCSH (10th edition) to determine if it matched any Library of Congress subject heading that was not included online in ORION. If it failed the single heading exact match category, she then used the browsing command to identify multiple heading matches. When no match was found, she searched the LCSH hard copy for matches to headings not available in ORION.

The results of this study showed that, when matching with user expression, single heading matches together with exact match, variation match, and partial match, accounted for 74% high. In other words, Library of Congress subject heading contributes significantly in subject keywords searching.

Jonathan C. Marner (1993) has a similar conclusion. He examined 425 bibliographic records from the Online Catalog of Texas A&M University Library to determine to what extent libraries can dispense with online cross-reference systems, assuming that keyword searches offer an adequate retrieval mechanism. He searched all of the defined headings (in the research) in the bibliographic records against local NOTIS authority file to retrieve their corresponding authority records. Most of the authority records obtained from this file were imported from OCLC Authority File without alteration. For non-matching records in NOTIS local authority file, Marner searched the OCLC Authority File. Marner then used the terms in 4xx "see from" field of each authority record to match with every variable field in the corresponding bibliographic record.

Marner found that matches resulting from 650/ 651 fields (topical and geographic subject headings) had the highest rate, a rate of 32.51%, when compared with matches resulting from the 100/110/111/130 (main entries) 16.27%; the 700/710/711/730 fields

(name/uniform title added entries) 14.03%; and the 245 field (title and statement of responsibility) 13.25%. His finding suggests that authority work and cross-references systems are of great value to an online catalog. He recommends that a typical search strategy be a keyword search to retrieve appropriate bibliographic records, and then use the terms in the authorized subject headings assigned to the records to conduct subject searches.

Ray R. Larson (1991), in his discussion on remedies to subject searching problems, discussed the following three major facets of online catalog system which need to be taken into consideration in improving subject searches:

1) The database;

2) The search processing and retrieval algorithms; and

3) The user interface.

Larson suggests that "no single method will provide a complete solution to the problems of subject searching, but each of the facets of the system need to be enhanced to contribute to a solution" (Larson, 213). Following studies look at search processing and system design.

Research Pertinent to System Design as an Approach to Improve Subject Access

Tschera Harness Connell (1991) conducted a study to determine system design that would help increase recall with data that already exists in current records. She randomly selected 1,023 titles from Book Review Digest and retrieved the titles' corresponding LC bibliographic records from OCLC. She then took a paragraph description of "what the book is about" from Book Review Digest to determine a match rate between the book description and the subject headings assigned, as well as keywords in the title proper of each of the 1,023 bibliographic records. The first phase of the study resulted in 35.7 terms or phrase exact matches on subfield a of the subject headings (6xx), and 3.6% on cross-references.

Contradicting Marner's findings (1993), Connell's findings did not support the assumption that LCSH "see" references will greatly increase recall (less than 4% were unmatched with book descriptions). Connell then matched the rest of book descriptions that had no match on subject headings, with the keywords in the title proper (field 245 subfield a), which resulted in 27.8% match.

The phase I of Connell's study demonstrated that the potential match rate for book description with main headings and title proper was 67%. To improve recall, Connell introduced five tests to match the remaining unmatched items with other segments of the bibliographic records. The test result indicated that matches of keywords subject field (between 37% to 47%) was greater than that of keyword title subfield (between 29% and 38%). Overall, the five approaches increased recall by 20%. Nearly 50% of the terms in the subject subfield, and about 27% of the terms in the title subfield represented "form" which would potentially retrieve large result sets.

Connell suggests that searching for keywords in the personal and corporate name fields as well as inverted headings and headings with parenthetical qualifiers, would increase precision more effectively.

Mary Micco and Rich Popp (1994) used an expert system, namely "Improving Library Subject Access (ILSA)" prototype with 100,000 MARC records, and 20,000 additional MARC records enhanced with table of contents terms to conduct their research. Their operational objective was to "link users' natural language terms to the controlled Library of Congress subject headings". The purpose of this study was to use the strengths of both natural language and controlled vocabulary to solve the problem of large retrieval sets.

They set a policy that the first heading assigned represented the "aboutness" of a document and is the primary heading. Due to the limited vocabulary of Library of Congress subject headings – most of them are general in nature, they further selected classification

number to cluster documents online because class number is being used to group items with similar subjects on the shelves. The subject clusters help manage human knowledge in a "tree" like hierarchy, which can be linked to provide a content for the term and enable broader or narrower a search query.

They indexed all keywords in the MARC records, and linked keywords to the subject clusters using automatic natural language scheme. The system instructed users from input terms to controlled vocabulary of the subject clusters, and consequently enabled users to narrow or broaden search queries with the adoption of hierarchical Dewey Classification numbers.

Additional keywords from the table of contents in the 20,000 MARC records increased 70 additional natural language terms per items, and resulted in the decrease of the number of zero-hit searches (Micco and Popp did not indicate the percentage of the zero-hit searches) to around 4%. The downside of this approach was that it also increased the size of retrieval sets, which further aggravated the large set problem. Therefore, Micco and Popp recommend the grouping of the subdivision assigned with the Library of Congress subject headings into larger clusters, which allows searchers to browse through huge retrieval sets to narrow down their search queries.

Micco and Popp found that taking the existing elements in the records and then manipulating them in a new way could improve subject access. The improved subject access outperformed the time and labor of the conversion effort.

Research Looking at Subject Access Analysis in Different Disciplines

Other researchers analyzed data for different disciplines. Raya Fidel (1992) observed 281 searches conducted by 47 professional searchers to determine whether controlled vocabulary or free-text was searched in each search session, and also to identify the reason

behind the choices. Fidel found that about half of the searchers chose controlled vocabulary, and the other half used free-text as search terms, depending on each searcher's preference and the specific situation of the search session. Fidel's analysis showed that searchers in Science and Technology group used free-text (76%) more frequently than other groups of searchers (Medicine 34%; Social Sciences and Humanities 39%). On the other hand, the data also indicated that all searchers (100%) in the Medicine group always checked a thesaurus before entering searches (Social Sciences and Humanities 87%; Science and Technology 68%). Fidel suggests that both free text (text words), and controlled vocabulary (descriptors) are indispensable for quality searching.

C.P.R. Dubois (1987) in his evaluation on issues of free text versus controlled vocabulary, identified semantics, context, relational structure, behavior, and discipline as the five major areas pertinent to the advantages and disadvantages of these two retrieval techniques. He commented "Some disciplines or areas of research are notably more rigid in their terminology than others. Moreover, some rapidly evolving areas may be extremely fluid in the terminology used to express the same concept" (Dubois, 247). He pointed out that chemical nomenclature, as an example, was notorious for its complexity. It used perhaps at least nine synonymous terms or codes for one single chemical. Law, on the other hand, may be a discipline that mainly had "unique and accepted" terms.

Monica Cahill McJunkin (1995) in her research on retrieval performance in terms of recall in title keyword searches, sampled items published from 1983-1985 in the subject areas of "Economics and Business". She compiled a list of Library of Congress subject headings assigned to the sample items from corresponding bibliographic records retrieved from the OCLC Online Union Catalog. The title keywords from the sample items were then searched with and without adjacency operators in the OCLC FirstSearch.

Her findings indicate that neither precision nor recall was high on title keyword searches with or without adjacency operators. She stated that "Many exact subject heading

matches were missed by title keyword searches" (McJunkin, 170). As the study tested items in Economics and Business, one can ask how well title keyword searches perform in other disciplines?

Carolyn O Frost (1989) analyzed 2,268 sample shelf-list cards extracted from the University of Michigan to find the percentage of bibliographic records where the controlled vocabulary of subject terms match the keyword-title. She matched title keywords with Library of Congress subject headings at "word" and "phrases" levels. All disciplines except literature were tested. Her study demonstrates that in exact match of entire heading, main heading keyword, and subdivision keyword, the title term in more than 53% of the bibliographic records analyzed found at least one subject heading term that matched it.

When looking at all levels of matches, including both exact and partial matches, Frost found that 73% of the sample contained word or words from the title that matched some part of the subject heading. Among the disciplines examined, the science and technology group had a 82% matching rate, the highest percentage of matches; humanities 74%, social sciences 72%; and history which has the lowest matching rate, a rate of 64%.

III. PROBLEM STATEMENT AND OBJECTIVES OF THE RESEARCH

This study considers whether the use of terms in title as searching vocabulary is an effective alternative to the use of controlled lists. When users have entered a word that appears in the title, to what extent this term will also appear as a subject heading, as part of a subject heading, or as a truncated part of a subject heading in each discipline? In other words, the general research question to be addressed by this study is to find the degree of matches that exist between the controlled vocabulary of subject headings and the terms in the title in different discipline. The hypotheses of this study are 1) sciences and technology subject areas have the highest match rate; 2) match rate in social sciences is much lower than that of sciences and technology subject areas; 3) title keyword is an effective alternative to subject heading in sciences and technology subject areas.

In reviewing studies in the area of improving subject access, some findings report that title keyword and subject searching complement each other (Peters and Kurth 1991; Tillotson 1995; Xu and Lancaster 1998). Others demonstrate that title keywords could easily retrieve relevant materials as does subject searching (Akeroyd 1990; Ensor 1992; Cherry 1992). Still others suggest that title keyword searching is good when used as a "lead in " to controlled vocabulary searching (Peters and Kurth 1991; Larson 1991; Frost 1989). Some indicate that controlled vocabulary searching using subject headings results in better precision and recall than do title keywords (Gerhan 1989; Carlyle 1989; Marner 1993). Different approaches have been explored to improve subject access. Some of these researchers collected data extracted from transaction logs (Peters and Kurth 1991; Akeroyd

1990; Larson 1991; Carlyle 1989), or bibliographic records from card catalogs (Carlyle 1989), or from online catalogs (Gerhan 1989; Frost 1989), and then analyzed data in all subject areas as a whole, not in a specific discipline, nor in each discipline. Only a few researchers have been devoted to the comparison of effectiveness of online subject access by discipline (Fidel 1992; McJunkin 1995; Frost 1989). To fill this gap in research, this study compares the effect of using keyword searching in title versus using subject headings among the 10 divisions of Dewey Decimal Classification.

The usefulness of information is a subjective matter, and the demand for precision and recall varies from one group of users to another. With reference to precision and recall, the demand of undergraduate students who search information for writing term papers will not be the same as that of doctoral students who search information for their dissertation. Since using title keywords for searching has become a feature of the online catalog, more users have switched from using subject heading index to title keywords when searching information. Larson (1991, 210) writes that "[t]he replacement of subject searching with title keyword searching indicated that users are attempting to avoid the search failure problems presented by LCSH"; and "[t]he switch to title keyword searching seems to indicate that the desire to do topical searching has not diminished, but that the penalties incurred by the user in the process of using the subject index have led to the decline in its use". Most library users do not understand LCSH well. As a result, they do not know how to effectively use the subject headings for finding the needed information. Allyson Carlyle (1989, 57) reported that "LCSH has long been regarded as a librarian's tool and not a general reference tool". Frost's study (1987) showed that only 40% of patrons responded that LCSH was the appropriate source of terms to use in online catalog subject searches (described by Cherry 1992, 95). If there is high percentage of match between title keywords and subject headings, library users may use title keywords to perform topical searching effectively, and catalogers may re-consider the needs of assigning subject

headings. This study focuses on word to word comparison between title keywords and subject heading terms among the 10 Dewey classification division to determine if title keyword an effective alternative to subject heading.

IV. METHODOLOGY

Books listed in the 1997, volume 6, number 2 issue of the OCLC Selected Titles for University and Research Libraries (OCLC Online Computer Library Center 1992-) are searched in WorldCat (the OCLC Online Union Catalog) via OCLC's online cataloging system. WorldCat is the largest database of bibliographic records in the world, and consists of 40 million unique records in 400 languages covering all subject matters.

OCLC selected Titles for University and Research Libraries is a tool for collection development. Books listed in the publication must have been selected and cataloged by 10 or more of the 121 research libraries. The books listed in the OCLC Selected Titles for University and Research Libraries are published in the current year or in the immediate past year, and cover all subject areas. The list is based on records entered into WorldCat (the OCLC Online Union Catalog) in the one-year period preceding each of the three quarterly issues. Titles are arranged by subject in the order of the 10 divisions of the Dewey Decimal Classification.

Every item listed in the selected issue of the OCLC Selected Titles for University and Research Libraries is searched in WorldCat for the corresponding record. A total of 923 entries were listed in the issue.

Many online catalogs support searching keywords in a title (subfield a, and/or subfield b). In addition, online catalogs can be designed to search just the main heading of a subject heading. The first phase of this study determines a match rate between the title proper (245 field, subfield a) as well as other title information (subfield b), and the first

element (subfield a) of Library Congress Subject Headings. All types of subject headings – personal name (600 field), corporate name (610 field), conference or meeting name (611 field), uniform title (630 field), topical (650 field), and geographic (651 field) – were compared. 14 stop words are excluded from comparison: a, an, and, at, by, for, from, how, in, of, on, the, to, with. The following matching criteria partially derived from Connell's 1991 study (page 91) were used to determine the level and degree of match of each record in each Dewey classification division.

For a term that occurs more than once in a record, it is always ranked according to its best match. For instance, if a term appears in two 6xx fields in a record, one in the first element of 6xx as a Subject Heading exact match, as well as a Keyword match in another 6xx field, the term is ranked as Subject Heading exact match.

A. Subject Heading Exact Match

A subject heading match is counted as an exact match when a term or a phrase in the title is exactly the same as a term or a phrase in the first element of a subject heading when compared from left to right, letter by letter (excluding capitalization, punctuation, and birth/death dates of persons). For a phrase match, the phrase in a title field has to match the term in the first element of a 6xx field in direct order. Subject heading exact match includes the following two categories by definition:

A. A "**Single Exact Match**" is counted when there is one Exact matched term or phrase in a record.

Example: Title (245) **Lunderston** tales
Subject (651) **Lunderston** (Scotland)

Title (245) Inorganic **materials**
Subject (650) **Materials**

Title (245) An A to Z of **feminist theology**
Subject (650) **Feminist theology**

B. A **"Multiple Subject Match"** is counted when there is more than one exact matched term(s), or phrase(s) in a record; it could be two or more exact term(s)/phrase(s) match, or one exact term/phrase match plus one or more keywords match.

Example: Title (245) **Security** challenges posed by **China**
Subject (651) **China**;
(650) National **Security**

(Note: The term "China" is an exact subject match, and the term "Security" is a keyword match. As a result, this entry is classified as "multiple subject match".)

Title (245) Work, leisure and well-being
Subject (650) **Work**;
(650) **Leisure**

(II). Keyword Match

A keyword match is counted when a term in the title is exactly the same as a term in the first element (or a term in parenthetical qualifier) of a subject heading when compared from left to right, letter by letter (excluding capitalization, punctuation, and birth/death dates of persons). Keyword Match includes the following three categories by definition:

A. **Multiple Keywords Match** is counted when there are more than one keyword match in a record, regardless the order.

Example: Title (245) Women's work and health in Britain
Subject (600) **Women's health** $z Great Britain
Title (245) Children and television
Subject (650) **Television** and **children**

B. **Keyword Plus Match** is counted when one exact matched keyword in the first element of 6xx fields plus a term that can be modified to become an exact matched term or phrase. Only following conditions are counted:

a. Truncation modification to a term on $ a of 6xx field

Example: Title (245) The changing *European* **security** environment
Subject (650) **Security**, International
(650) *Europe* $x Defenses

b. Plural form modification to a term on $a of a 6xx field

Example: Title (245) Why vote **Liberal** *Democrat*?
Subject (650) **Liberal** *Democrats*
Title (245) Basic principles of *membrane* **technology**
Subject (650) *Membranes* (**Technology**)

c. Exact term in subdivisions (such as $z, $x, or $p) of a 6xx field

Examples: Title (245) From self-help housing to sustainable settlement: capitalist development and urban planning in *Lusaka, Zambia*
Subject (650) Housing policy $z *Zambia*, $z *Lusaka*

d. Acronym of a term on $a of a 6xx field

Example: Title (245) *MHC* molecules and antigen processing
Subject (650) *Major Histocompatibility Complex* $x physiology

C. **Single Keyword Match** is counted when there is only one keyword match in the first element of a 6xx field.

Example:	Title (245)	Life on the **Mississippi**
	Subject (651)	**Mississippi** River Valley

(III). No match.

The second phase of this study deals with the remaining unmatched records as there are some cases where a term in the first element of a 6xx fields can be modified to become an exact matched term. The second phase of the study determines to what extent words in the title (including subfields a and b) match part of the word(s) in the first elements of the main subject fields. Because headings in subdivisions in 6xx fields can not stand alone, subdivisions in 6xx fields were not tested.

All remaining unmatched items were examined. WorldCat, as well as some online catalogs provide word truncation capability in keyword searching, which allows users to mask one or more than one characters in a keyword search string. Using this approach, keywords that are in plural form, variant in spelling, or variant in suffix can be retrieved. The result of the test helps determine, when considering just recall, the maximum match rate between title keyword and subject headings with the help of appropriate system design. Examples are given as follows:

A. Plural form

Example:	Title (245)	**Probability** theory: collection of problems

	Subject (650)	**Probabilities** $x problems, exercises, etc.

In WorldCat (OCLC Online Union Catalog), the system supports character(s) masking. For example:

Keyword	Retrieves
Adverti#e	advertize, advertise
Wom#n	woman, women

B. Variation in suffix

Example:	Title (245)	**Placental** pharmacology
	Subject (650)	**Placenta** $x Metabolism

In WorldCat, the system supports both truncation and wild cards. For example:

Keyword	Retrieves
Computer?	computer, computerization, computerrized, computers
Librar?	librarian, librarians, librarianship, libraries, Librar

C. Acronym

Example:	Title (245)	Mixed **IC** design
	Subject (650)	**Integrated circuits** $x design and construction

In other words, through system design, more relevant information can be retrieved, and thus increase the rate of recall.

V. RESULTS

From among 923 titles from OCLC Selected Titles for University and Research Libraries, volume 6, number 2, published in July, 1997, 907 records were analyzed. Excluded were foreign titles, and records which had no subject headings assigned. In 42.78% (388/907) of the records analyzed, keywords in the titles exactly matched a complete subject heading in the first element (subfield a) of 6xx. In terms of disciplines based on Dewey classification divisions, the range of match was from 19.04% to 56.20%.

The percentage of "subject heading exact match" is the most important indicator when considering whether title keywords could be an alternative to subject headings. The reasons are that: 1) the function of a main subject heading is to provide access by subject to all relevant materials in a given collection, and 2) the title term(s) in the "subject heading exact match" category is identical to the term(s) in the first element of the subject heading field.

However, multiple keywords or single keyword are also likely to retrieve some useful information. Bates (1989) states that "[i]n online catalogs, title keyword searching can constitute a powerful kind of subject searching. Keyword matching with one or two title words - either words from a known title or 'just fishing' - can often produce a number of highly relevant titles" (Bates, 403). Frost (1989) writes that "[t]he retrieval value of a match on the main heading keyword can vary, depending on the distinctiveness and the number of the matching words" (Frost, 173). On the other hand, as McJunken (1995)

pointed out, "[s]ingle keywords tended to be general terms that resulted in large retrieval set" (McJunken, 169). As a result, in an attempt to look at possible matches through multiple title keywords, the secondary match rate obtained in the present study was to combine matches resulting from multiple keywords match, and single keyword plus match with matches resulting from subject exact match. The last match rate examined results from combining the figure with matches resulting from partial matches. Following is the result of the above analysis in each Dewey division:

In the 000 division, subject exact match (meaning title keyword(s) match exactly the same as term(s) in the first element of 6xx field) accounts for 47.13% (41/87) of the total records analyzed. Combined with matches resulting from multiple keyword match (10.34%; 9/87), and keyword plus match (5.75%; 5/87) the figure is 63.22%. Combining this figure with matches resulting from partial match (2.30%; 2/87) in the first elements of subject heading fields the match rate is 65.52%. Table one is a summary of the analysis.

Table 1.--Analysis of 000 Division

000 Generalities Division (n=87)		number of records	Total percentage in the division
Subject Exact Match:		41	47.13%
Single exact match:25; Multiple exact match:13			
Keyword Match:		29	33.33%
Multi KW:	9		
Single KW Plus:	6		
Single KW:	15		
No Match:		17	19.54%
Partial match in the main element of 6xx:	2		

Similar analysis for divisions 100-900 are represented in tables 2-10.

Table 2.--Analysis of 100 Division

100 Philosophy & Psychology Division (n=32)		number of records	Total percentage in the division
Subject Exact Match:		15	46.88%
Single exact match:7; Multiple exact match:8			
Keyword Match:		9	28.12%
Multi KW:	4		
Single KW Plus:	0		
Single KW:	5		
No Match:		8	25.00%
Partial match in the main element of 6xx: 3			

Table 3.--Analysis of 200 Division

200 Religion Division (n=29)		number of records	Total percentage in the division
Subject Exact Match:		10	34.48%
Single exact match: 7; Multiple exact match: 3			
Keyword Match:		9	31.04%
Multi KW:	5		
Single KW Plus:	2		
Single KW:	2		
No Match:		10	34.48%
Partial match in the main element of 6xx: 3			

Table 4.--Analysis of 300 Division

300 Social Sciences Division (n=261)		number of records	Total percentage in the division
Subject Exact Match:		108	41.38%
Single exact match:64; Multiple exact match:44			
Keyword Match:		81	31.03%
Multi KW:	32		
Single KW Plus:	24		
Single KW:	25		
No Match:		72	27.59%
Partial match in the main element of 6xx: 28			

Table 5.--Analysis of 400 Division

400 language Division (n=7)		number of records	Total percentage in the division
Subject Exact Match:		3	42.86%
Single exact match:1; Multiple exact match:2			
Keyword Match:		1	14.28%
Multi KW:	0		
Single KW Plus:	0		
Single KW:	1		
No Match:		3	42.86%
Partial match in the main element of 6xx:	3		

Table 6.--Analysis of 500 Division

500 Natural Sciences & Mathematics Division (n=87)		number of records	Total percentage in the division
Subject Exact Match:		77	56.20%
Single exact match:43; Multiple exact match:34			
Keyword Match:		39	28.47%
Multi KW:	16		
Single KW Plus:	5		
Single KW:	18		
No Match:		21	15.33%
Partial match in the main element of 6xx:	13		

Table 7.--Analysis of 600 Division

600 Technology (Applied Sciences) Division (n=221)		number of records	Total percentage in the division
Subject Exact Match:		92	41.63%
Single exact match:51; Multiple exact match:41			
Keyword Match:		70	31.67%
Multi KW:	29		
Single KW Plus:	17		
Single KW:	24		
No Match:		59	26.70%
Partial match in the main element of 6xx:	2		

Table 8.--Analysis of 700 Division

700 The Arts Division (n=49)		number of records	Total percentage in the division
Subject Exact Match:		11	22.45%
Single exact match:5; Multiple exact match:6			
Keyword Match:		26	53.06%
Multi KW:	7		
Single KW Plus:	1		
Single KW:	18		
No Match:		12	24.49%
Partial match in the main element of 6xx:	5		

Table 9.--Analysis of 800 Division

800 Literature & Rhetoric Division (n=42)		number of records	Total percentage in the division
Subject Exact Match:		8	19.04%
Single exact match:6; Multiple exact match:2			
Keyword Match:		17	40.48%
Multi KW:	7		
Single KW Plus:	1		
Single KW:	9		
No Match:		17	40.48%
Partial match in the main element of 6xx:	2		

Table 10.--Analysis of 900 Division

900 Geography & History Division (n=42)		number of records	Total percentage in the division
Subject Exact Match:		23	54.76%
Single exact match:18; Multiple exact match:5			
Keyword Match:		12	28.57%
Multi KW:	9		
Single KW Plus:	1		
Single KW:	2		
No Match:		7	16.67%
Partial match in the main element of 6xx:	1		

VI. DISCUSSION

It should be noted that this study examines the entire population of the 2nd issue of OCLC Selected Titles for University and Research Libraries for July, 1997. Because the entire population was examined, no inferential statistical testing was performed in the comparative analysis below. The author realizes that these results may not hold the true for other populations, for example for items in the 2nd issue of OCLC Selected Titles for University and Research Libraries for July, 1998.

Among the 10 Dewey divisions, the 500 division, disciplines in natural sciences and mathematics, accounts for the highest percentage of subject heading exact match, 56.20%. Therefore, first hypothesis that, sciences and technology subject areas have the highest match rate, does not hold true for this population. The 900 division, disciplines in geography and history, accounts for the second highest percentage of subject heading exact match, 54.76%. Within the 900s, 38.10% (16/42) of the records counted as subject heading exact match are from 651 geographical subject heading field.

On the other hand, the 800 division, disciplines in literature and rhetoric, accounts for the lowest percentage of subject exact match, 19.04%. Many titles in the 800 division carry personal names, which match the 600 personal name subject headings. However, the matches are counted as multiple (two) keywords match, not as subject exact match, according to the matching definition. When combined the figure with matches resulting from multiple keywords match in the 600 field (14.29%; 6/42), it increases the percentage

from 19.04% to 33.34%. The 700 division, disciplines in arts, accounts for the second lowest percentage of subject exact match, 22.45%.

One thing that is worth mentioning is the 300 division, disciplines in social sciences. 41.38% of the records in the 300 division account for subject exact match. In the 600 division, disciplines in technology (applied sciences), 41.63% of the records account for subject exact match, which is just slightly higher than the match rate of the 300 division. The findings reject the second hypothesis that, match rate in social sciences is much lower than that of sciences and technology subject areas.

When looking at keyword match, including the total matches resulting from multiple keywords match, single keyword plus match, and single keyword, in the present study, the 700 division ranked the highest level of match: 53.06%, followed by the 800 division: 40.48%. The lowest level of keyword match is in the 400 division: 14.28%, followed by the 100 division: 28.12%.

When looking at the percentage of "no match" category, two Dewey divisions' "no match" rates are very close -- 42.86% in the 400 division, and 40.48% in the 800 division. On the other hand, when ruling out the percentage of "partial match" from the "no match" set, the highest "no match" rate is 35.71% in the 800 division.

Table 11 is a summary of levels of match between keywords in titles and subject headings in the 10 Dewey Divisions.

Table 11.--Levels of Match between Keywords in Titles and Subject Headings in the 10 Dewey Divisions

DDC Divisions	Exact Match	Keyword Match	No Match
000 n= 87	41 (47.13%)	29 (33.33%)	17 (19.54%)
100 n= 32	15 (46.88)	9 (28.12)	8 (25.00)
200 n= 29	10 (34.48)	9 (31.04)	10 (34.48)

Table 11.--Levels of Match between Keywords in Titles and Subject Headings in the 10 Dewey Divisions –continued

DDC Divisions	Exact Match	Keyword Match	No Match
300 n=261	108 (41.38%)	81 (31.03%)	72 (27.59%)
400 n= 7	3 (42.86)	1 (14.28)	3 (42.86)
500 n= 87	77 (56.20)	39 (28.47)	21 (15.33)
600 n=221	92 (41.63)	70 (31.67)	59 (26.70)
700 n= 49	11 (22.45)	26 (53.06)	12 (24.49)
800 n= 42	8 (19.04)	17 (40.48)	17 (40.48)
900 n= 42	23 (54.76)	12 (28.57)	7 (16.67)

In short, the findings indicate that title keyword is not a legitimate alternative to subject heading, as the matches from subject exact match, ranging from 19.04% low to 56.20% high. Therefore, the third hypothesis that, title keyword is an effective alternative to subject heading in sciences and technology subject areas, does not hold true for the above population.

The findings of the study are different from that of Frost's study (1989). Frost analyzed 2,268 records in all disciplines based on Library of Congress Classification. Frost defined six categories of analysis. Among which, the combination of Frost's exact match - entire heading, and exact match - main heading component of subdivided heading - is similar with subject exact mach of the present study. Frost's findings show that, in the combination of matches on entire heading and matches on main heading, only 23% of records fall into the category versus 42.78% of the present study. It is not legitimate to compare the two findings by discipline, because one is based on LC classification, and the other is based on Dewey classification. Nevertheless, Frost reports 33% (20% from matches on entire heading plus 13% from matches on main heading) subject exact match

rate in disciplines of science and technology, 24% in social sciences, 9% in humanities, and 16% in history. Generally speaking, Frost's match rates are lower than that of the present study.

Hong Xu and Lancaster (1998) analyzed 205 items selected from materials classified in Dewey classification classes 300, 500, 600, and 700. Xu took a different approach by assigning 844 unique subject access points (SAPs) to the 205 items. Xu's findings show that 30.03% overlap among the SAPs in both titles and subject headings. The methodology and research design of Xu's study are not the same as that of the present study. However, when compared the subject heading exact match figures from 300, 500, 600, and 700 of the present study, it shows only 4.74% difference between the two findings (34.77% in the present study, and 30.03% in Xu's study).

VII. CONCLUSION

The results of the study show that title keyword is not an effective alternative to subject heading, as the match rate of subject heading exact match among the 10 Dewey classification divisions, on average, is 42.78% (388/907). This is less than half which does not justify title keyword searching as an alternative. When looking at each discipline, the 500 natural sciences and mathematics division had the highest subject heading exact match: 56.20%, which is still too low to justify title keyword searching as an effective alternative in the given subject areas.

The average figure of keyword match (32.30%; 293/907) is of limited value to the study, because the chances of obtaining large result sets via keyword searching are potentially high. Although several studies have suggested approaches of solving large result sets, such as through the effort of word stemming, truncation, etc., user satisfaction could be still low. The point has been clearly stated by Tillotson that "[s]ome keyword searches provided citations that appeared to be about the topic but were still declared unsuccessful by the searcher" (Tillotson 1995, 203). Naturally, for search queries that generate zero-hit, keyword searching would play a significant role in finding something that might be useful to users, because 32.30% of the records contain at least one keyword that is not available through subject headings. When the keyword searching failed, the partial match approach could come into place in solving the problem of zero-hit queries. However, the percentage in the partial match category is only 6.84% low (62/907).

Findings of the study also indicate that people's perception toward title keyword is not the same as the reality, that is that, even in disciplines of science and technology, the average match rate: 41.63% between title keyword and subject heading is not high. And, the match rate in disciplines of social sciences: 41.38% is very close to that of science and technology. The data support the need for assigning subject headings in every discipline, and that the assignment of subject headings is still indispensable when cataloging materials.

The strengths and weakness of title keyword and subject headings are not the focuses of the present study, as there have been many studies that have dealt with the merits of controlled language and uncontrolled language. In addition, there have been many studies that have devoted to the enhancement and improvement of the Library of Congress Subject Headings, which are beyond the scope of the present study. The data of the present study do show that title keyword does contain subject related natural language, as 75.08% (42.78% plus 32.30%) of the records contain at least one single keyword that matches a term in the first element of the subject heading field, at least for this population. For users who are not familiar with Library of Congress subject headings, title keyword indeed could serve as a "lead-in" to subject headings. As Peters and Kurth (1991) suggest, title keyword is not primarily an option of last resort in subject searching and that a bridge that allows users to go from items retrieved by keyword to other bibliographic records containing the same subject headings would be a useful enhancement (described by Arlene G. Taylor 1992, 317). Given the result of this study, the following approach might improve end-users title subject searching to a greater degree: 1) establish a link between title keywords and bibliographic records containing the same subject headings, 2) then link the headings to an online subject heading thesauri, which allows users who are not certain if the headings are appropriate to choose appropriate headings from the browsing index of the structured subject headings, 3) and provide "hot link" to each heading in the thesauri back

to bibliographic records containing the heading. These suggestions are ideas for further study.

REFERENCE LIST

Akeroyd, John. 1990. Information seeking in online catalogs. Journal of Documentation 46, no. 1 (1990): 33-52.

Bates, M.J. 1989. Re-thinking subject catalouging in the online environment. Library Resources and Technical Services 33 (October): 403-11.

Carlyle, Allyson. 1989. Matching LCSH and user vocabulary in the library catalog. Cataloging & Classification Quarterly 10, no.1-2 (1989): 37-63.

Cherry, Joan M. 1992. Improving subject access in OPACs: an exploratory study of conversion of users' queries. Journal of Academic Librarianship 18, no. 2 (1992): 95-99.

Connell, Tschera Harkness. 1991. Techniques to improve subject retrieval in online catalogs: flexible access to elements in the bibliographic record. Information Technology and Libraries 10 (June): 87-98.

Dubois, C.P.R. 1987 Free text versus controlled vocabulary: a reassessment. Online Review 11 (August): 243-53.

Ensor, Pat. 1992. Users characteristics of keyword searching in an OPAC. College & Research Libraries 53 (January): 72-80.

Fidel, Raya. 1992. Who needs controlled vocabulary? Special Libraries 83, no. 1 (Winter):1-9.

Frost, Carolyn O. 1989. Title words as entry vocabulary to LCSH: correlation between assigned LCSH terms and derived terms from titles in bibliographic records with implementations for subject access in online catalogs. Cataloging & Classification Quarterly 10, no1-2 (1989): 165-79.

---1987. Subject searching in an online catalog. Information Technology Libraries 6 (March): 60-63.

Gerhan, David. R. 1989. LCSH in vivo: subject searching performance and strategy in the OPAC era. Journal of Academic Librarianship 15, no. 2 (1989): 83-89.

Larson, Ray. 1991. The decline of subject searching: long term trends and patterns of index use in an online catalog. Journal of the American Society for Information Science 42, no. 3 (1991): 197-215.

Marner, Jonathan C. 1993. Measuring the success of keyword search strategy in an online catalog. Technical Services Quarterly 11, no. 2 (1993): 1-11

McJunkin, Monica Cahill. 1995. Precision and recall in title keyword searches. Information Technology and Libraries 14 (September): 161-71.

Micco, Mary, and Popp, Rich. 1994. Improving library subject access (ILSA): a theory of clustering based in classification. Library Hi Tech 12, no. 1 (1994): 55-66.

OCLC Online Computer Library Center. 1992-. OCLC Selected Titles for University and Research Libraries. Vol 6, no. 6 (Dublin, OH: OCLC Online Computer Library Center, Inc., 1997)

Peters, Thomas A., and Kurth, Martine. 1991. Controlled and uncontrolled vocabulary subject searching in an Academic Library online catalog. Information Technology and Libraries 10 (September): 201-11.

Taylor, Arlene G. 1992. Enhancing subject access in online systems: the year's work in subject analysis, 1991. Library Resources & Technical Services 36 (July): 316-32.

Tillotson, Joy. 1995. Is keyword searching the answer? College & Research Libraries 56 (May): 199-206.

Xu, Hong, and Lancaster, F.W. 1998. Redundancy and uniqueness of subject access points in online catalogs. Library Resources and Technical Services 42, no. 1(1990): 61-66.